

Primeras pistas de estudio computacional de la recepción del *Regimiento de los Príncipes*.

Las marcas de lectura del manuscrito K.I.5 (Biblioteca del Escorial).

Journée d'études « Cartographies ibériques du livre médiéval »

Matthias GILLE LEVENSON

École nationale des chartes – CJM / École Normale Supérieure de Lyon - CIHAM

prénom [point] gille [tiret] levenson [at] ens-lyon [point] org

15 mars 2024



1 Introducción

2 Métodos de extracción de los fragmentos anotados

- Identificación de las marcas de lectura
- Segmentación de los fragmentos identificados
- Transcripción de los fragmentos segmentados

3 Estudio lexicométrico y primeros resultados

- Lectura distante : comparación de las frecuencias relativas de algunos términos
- Distribución de las marcas en el manuscrito
- Orientaciones de la lectura : algunos ejemplos

4 Conclusiones

1 Introducción

2 Métodos de extracción de los fragmentos anotados

3 Estudio lexicométrico y primeros resultados

4 Conclusiones

El *Regimiento de los príncipes*

- Traducción del *De Reginime Principum* de Egidio Romano, escrito hacia 1276-80
- Difusión de la tripartición de la filosofía práctica (moral, economía, política)
- Supuestamente traducido para la educación de Pedro I a encargo del entorno de Alfonso 11
- Una tradición con más huellas de lectura que otras, como la francesa (PERRET 2011).
- Enlace a los datos y programas :
https://github.com/matgille/extraccion_marcas_lecturas

La versión B

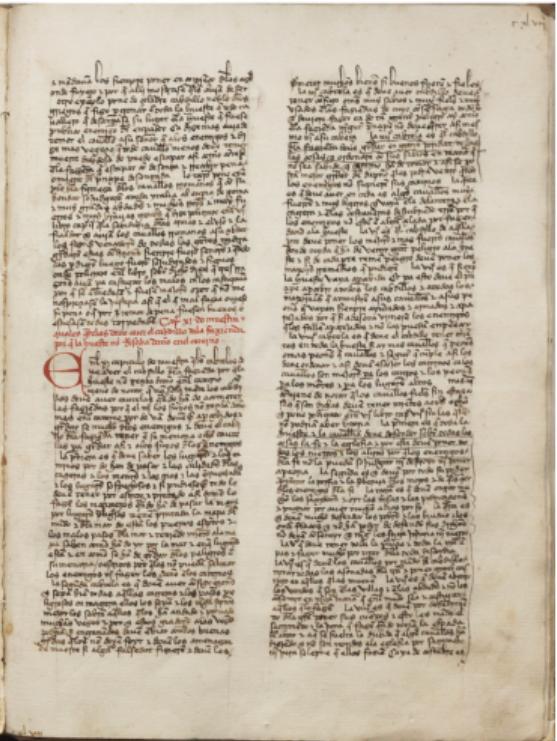
- Existen varias versiones del texto (A, B, C) (ACERO-DURÁNTEZ, DÍEZ GARRETAS et FRADEJAS RUEDA 2005), y circulan esencialmente manuscritos con una glosa extensa
- Aunque hay manuscritos sin glosa que corresponden, creo, a la versión original que debió existir sin las glosas (GILLE LEVENSON 2023)

El manuscrito K.I.5 del Escorial

Imagen del manuscrito

- Pertenece a la versión B, con glosas y una traducción generalmente resumida
- Un manuscrito muy peculiar en cuanto a sus características externas
- Una gótica cursiva probablemente del siglo 15.
- marcas frecuentes, sin glosas marginales o interlineales

El manuscrito K.I.5 del Escorial



K.I.5 f. 162v
Las marcas de lectura del mss. K.I.5

Tipos de marcas del K.I.5

- Dos tipos de marcas verticales (a y b)
- Marcas horizontales : el texto viene subrayado con lápiz.

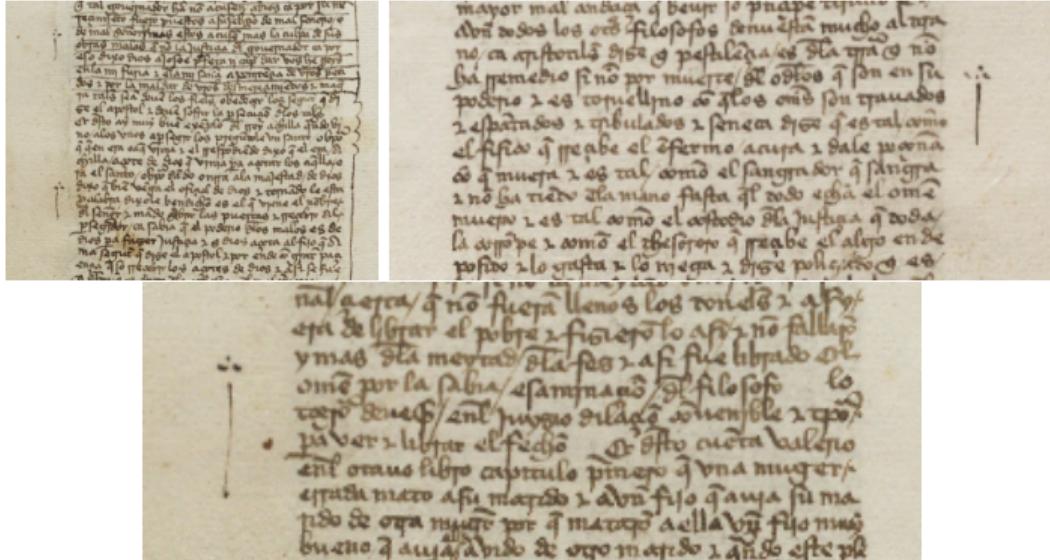
Marcas verticales : tipo a



Marcas verticales : tipo a



Marcas verticales : tipo b



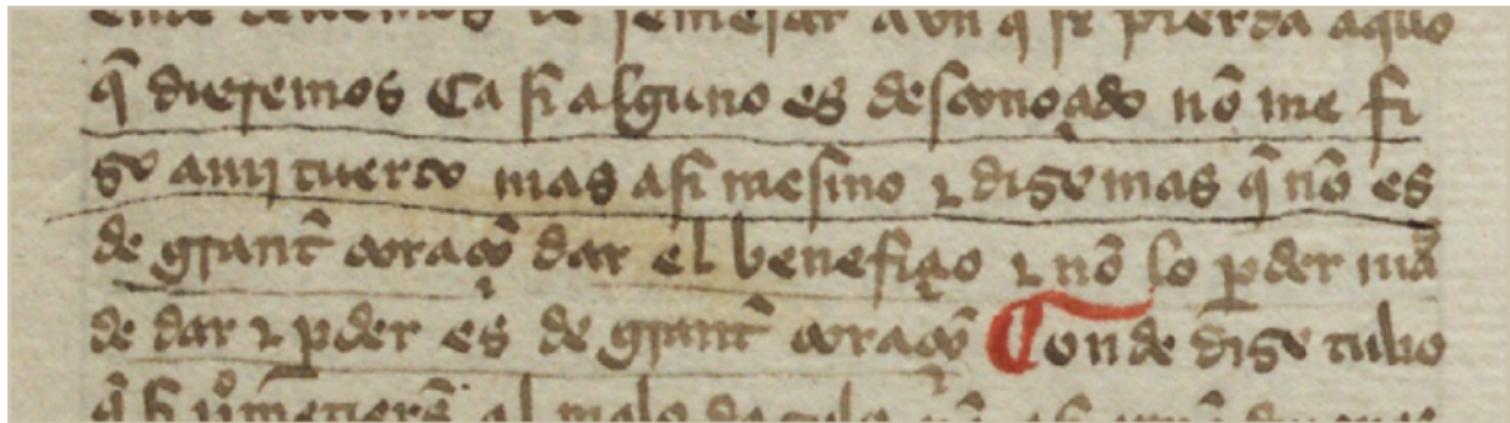
Marcas horizontales

mēto q̄ fasía nūca viuā dīcta el tēplo onde
 julho cesar ve so q̄ nō tomasē nada del templo m̄
 guno nō lo despo) ase sopena de muerte 2 alq̄ q̄
 se q̄la gueste de bren o nūca fue venada farta
 G amero el tēplo de apolone 2 lo despo) q̄so
 mēsmo dīse poliquato en llibro sobr̄ dīcho q̄dīse
 G pimpeyo nūca confirma q̄ despo) ase tēplo m̄ q̄
 q̄so mēsmo dīse alixadte q̄ p̄dona los tēplos
 2 allos q̄ estauā en ellos 2 si esto fasía los caud
 illos queríalos mucho mas los caudillos vestian
 seu e oniar la eglefia Conf xñ do mnestra en q̄

Marcas horizontales

En los sabios / onde aceta poligato en lus li
 bro q̄ nycenya biē fue éla amnidad de Dña
 no se acuerda q̄ los p̄nages homanoz fuesen
 sin leqaz sin sabiduria y despues q̄ fallecio la
 virna señac leqaz en los p̄nages fue enferma
 en los la mano dela caualla armada y fue cor
 ta la raya q̄ su p̄naga dgo y senorio y esto no
 sin fago calos p̄nages no puede mucho durar
 en su senorio sin sabiduria e onde el Rey d. Ro
 ma q̄bio desir al Rey de frana q̄figiese appre
 der leqaz atodos sus fyoz alegaz le el dicho
 d. poligato q̄ el Rey sin leqaz sin sabiduria
 en tal como el asno aronado **S**el por ende los

Marcas horizontales



... jemejar aun q se puerda aquo
q diezemos. Ca si afgano es desconocido no me fi
so a mi tuerco mas asi mesimo l dizemas q no es
de gran corazon dar el beneficio l no lo pderia
de dar l pder es de gran corazon. **Conde** dice tulio
a su imitero al male o mala l de q no se que

Tipos de marcas del K.I.5

- No conocemos la identidad de la persona que anota, pero las marcas verticales de tipo a y los subrayados son de la misma mano ; en general, ambas se completan.
- Las marcas verticales suelen cubrir mucho más texto.
- En cuanto a las marcas de tipo b, bastante frecuentes, es más difícil saberlo precisamente.

Orientaciones de estudio

- ¿Qué nos dicen las marcas de lectura de los intereses de un lector del 15 o del 16 ?
- ¿Podemos estudiar las marcas de manera exhaustiva, recuperando el texto anotado de manera semi-automática ?
- ¿Podemos identificar un programa de lectura preciso e intereses particulares ?
- ¿O es una lectura más general, centrándose en los elementos fundamentales de cada capítulo ?

1 Introducción

2 Métodos de extracción de los fragmentos anotados

- Identificación de las marcas de lectura
- Segmentación de los fragmentos identificados
- Transcripción de los fragmentos segmentados

3 Estudio lexicométrico y primeros resultados

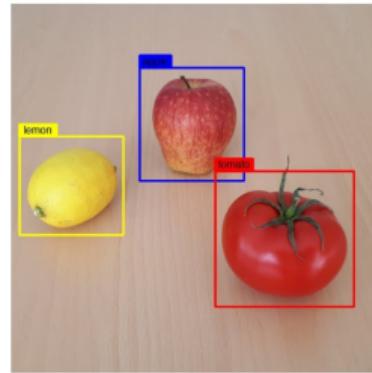
4 Conclusiones

Tecnología : el aprendizaje supervisado

- Las tres fases de la extracción hacen uso de una rama de la « inteligencia artificial » : el **aprendizaje supervisado**
- Se trata de aprender a la máquina a resolver un problema dándole parejas de **{datos + etiquetas}**

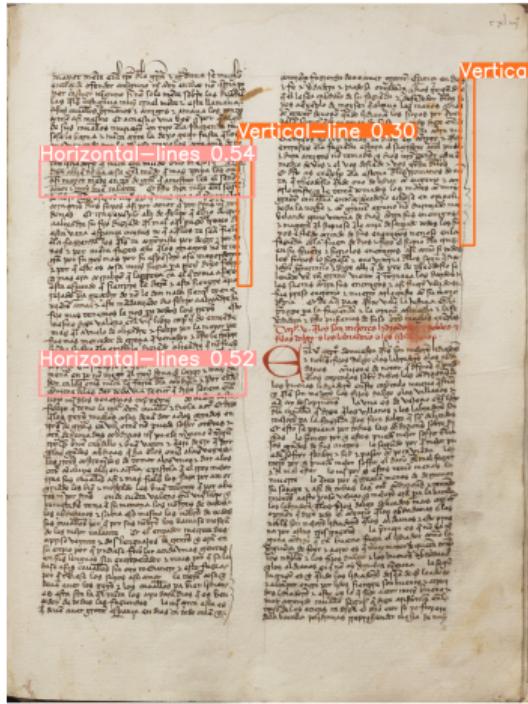
Herramienta y modelo de extracción de las marcas

- Se utiliza un método de **detección de objetos** con la herramienta YOLOv8 (REDMON, DIVVALA, GIRSHICK et FARHADI 2016).



- Esta herramienta ya se ha utilizado para identificar zonas de texto como columnas o miniaturas (CLÉRICE 2022).
- El modelo base se afina con 220 imágenes, o sea los 2/3 del corpus, lo que no es nada eficiente ; sin embargo, el modelo se podrá reutilizar luego para ampliar el estudio.

Un ejemplo de resultado



Resultados

- Los resultados son bastante buenos, pero con los dos tercios del corpus transcritos a mano, y el último segmentado automáticamente
- 153 marcas tras corrección.
- El modelo identifica correctamente todas las marcas pero hay algunas identificaciones múltiples

Herramienta y modelo de segmentación utilizado

- El motor de segmentación Kraken (KIESSLING 2019) se utiliza.

Un ejemplo de resultado de segmentación



Herramienta y modelo de transcripción utilizado

- Se utiliza el motor de transcripción Kraken (KIESSLING 2019) y el modelo CATMuS como base : 10 lenguas, siglos 7-14, corpus de entrenamiento de 160.000 líneas (CLÉRICE et al. 2024).

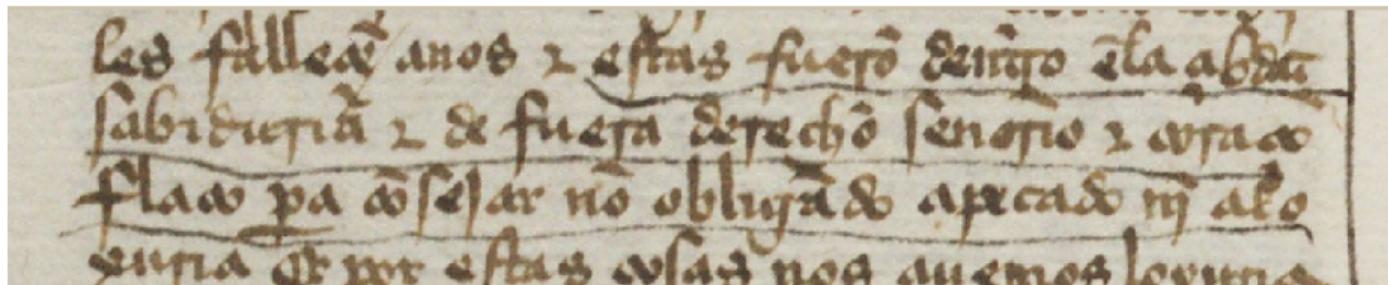


- El juego de datos de CATMuS tiene muy poca cursiva, y se tiene que adaptar el modelo a las escrituras del manuscrito con unas **1500 líneas**, en parte sin marcas, en parte con marcas
- Dos problemas :
 - 1 La gótica cursiva en sí
 - 2 Las propias marcas que vienen dificultar la lectura

Resultados

- Los resultados son un poco decepcionantes para una lectura fina del texto, pero sirven para lectura distante
- Esto se debe claramente a las marcas de lectura : el modelo es mucho mejor con el texto sin marcas, y los errores de segmentación y transcripción se combinan.

Ejemplo de transcripción de buena calidad



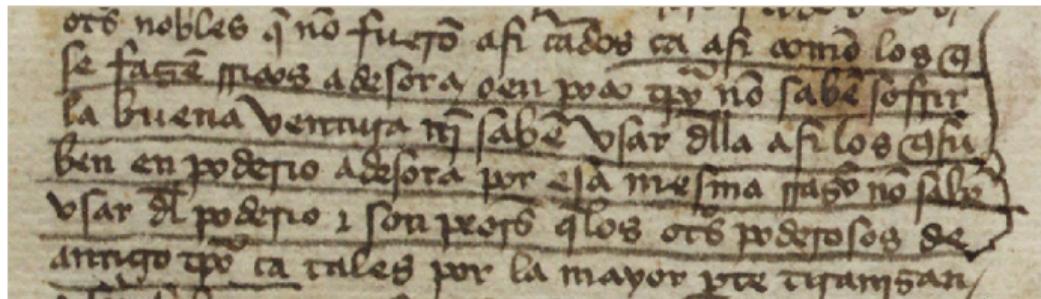
les falleçē anos e estas fuerō dentro ëla çibdar
sabiduriā e de fuerā derecho senorio e coraço
flaco pa cõsejar nō obligādo aipecado nī alo
xūria et por estas cosas nos auemos loxuria

les falleçē anos e estas fuerō dentro ëla çibdar
sabiduriā e de fuerā derecho senorio e coraço
flaco pa cõsejar nō obligādo aipecado nī alo
xūria et por estas cosas nos auemos loxuria

Una vez normalizado :

les fallecen a nos e estas fueron dentro enla cibdar sabidurian e de fueran derecho senorio e coraço
flaco para consejar non obligando ai pecado nin aloxunria et por estas cosas nos auemos loxuria

Ejemplo de transcripción de mala calidad



ots nobles q̄ nō fuerō así cādos ca así como los q̄
efaciē rriçes adesora oen poco tpx nō sabē sofrir
a buena ventura
bē
vsar dlla así los al
den en poderio adesora por esa mesma rraro nō sabe
osar dlt poderio e son peorō q̄los ots poderosos de
antigo tpo ca tales por la mayor pte tiranizan/

1 Introducción

2 Métodos de extracción de los fragmentos anotados

3 Estudio lexicométrico y primeros resultados

- Lectura distante : comparación de las frecuencias relativas de algunos términos
- Distribución de las marcas en el manuscrito
- Orientaciones de la lectura : algunos ejemplos

4 Conclusiones

Lectura distante, lectura atenta

- Franco MORETTI. *Distant Reading*. Verso, 2013
- La lexicometría como una forma de lectura distante, global, de un texto

Metodología

- Transcripción del texto de cada anotación
- Normalización y desarrollo de abreviaturas
- Lematización con Freeling
- Recuperación de las frecuencias y comparación con las frecuencias del texto completo **lematizado**
(base : incunable, edición de 1494 de Ungut y Polono)

Estadísticas básicas del corpus

- 360.000 palabras en el texto completo (base incunable)
- 10.000 palabras recuperadas en las marcas

Comparación de frecuencias

- La idea es conseguir las **frecuencias relativas** de cada lema (= número de ocurrencias dividido por el tamaño del corpus), y comparar estas frecuencias con las frecuencias del texto completo, para identificar alguna especificidad léxica

Métricas utilizadas

- Frecuencia absoluta (lemas) : número de ocurrencia de un lema en el corpus
- Frecuencia relativa : $\frac{\text{Frecuencia absoluta (lemas)}}{\text{tamaño del corpus}}$
- Frecuencia comparada : $\frac{\text{Frecuencia relativa (lemas) texto marcas}}{\text{Frecuencia relativa (lemas) texto completo}}$
- Si la frecuencia comparada supera 1, el lema aparece – en relativo – más en los fragmentos anotados que en el texto completo, lo que puede significar un interés superior por el término

Términos morales

- Un interés menor por las virtudes, aunque no inexistente

Lemma	Frec. abs. (fragm)	Frec. rel (fragm)	Frec. abs. (incunable)	Razón
virtud	21	0,00217	0,00175	1,237
fortaleza	3	0,00031	0,00043	0,71939
prudencia	1	0,0001	0,00017	0,59318
piedad	1	0,0001	0,00017	0,60377
soberbia	1	0,0001	0,00021	0,49722
lujuria	1	0,0001	0,00019	0,53669
misericordia	1	0,0001	0,00016	0,62613

Economía y gestión de la casa

- Poco interés por cuestiones de economía, desarrolladas en la segunda parte :

Lemma	Frec. abs. (fragm)	Frec. rel (fragm)	Frec. abs. (incunable)	Razón
mujer	9	0,00093	0,00193	0,48073
marido	1	0,0001	0,00053	0,19658
siervo	3	0,00031	0,00087	0,35716
hijo	17	0,00176	0,00221	0,79501
padre	7	0,00072	0,00126	0,57586
castigo	1	0,0001	0,00028	0,36356
casa	3	0,00031	0,00108	0,28654
dinero	1	0,0001	0,00039	0,2621

Términos políticos

- Lo que interesa el lector parece ser el actor político como individuo, sea bueno o malo

Lemma	Frec. abs. (fragm)	Frec. rel (fragm)	Frec. abs. (incunable)	Razón
príncipe	32	0,0033	0,00128	2,58224
rey	60	0,0062	0,0043	1,44184
tirano	12	0,00124	0,0004	3,12104
poderío	10	0,00103	0,00053	1,93207
caballero	21	0,00217	0,00117	1,85388
noble	18	0,00186	0,00071	2,61203
ciudad	14	0,00145	0,00177	0,81754
reino	11	0,00114	0,00132	0,86093

La tiranía

- La definición de la tiranía y de sus características es puesta de relieve, en 10 fragmentos distintos casi todos localizados en III,2 (Del Gobierno de la ciudad en tiempos de paz)
- Servir a Dios (frag. 242_2), buscar el bien común (90_0, 234_0), guardar la justicia (240_0), defender y conservar el patrimonio del Reino (241_0) como elementos de distinción entre tirano y monarca, así como algunos ejemplos (el tirano como Simia sobre el trono, según san Bernardo : 246_1) que muestran un interés particular por la tiranía.

Nobles y nobleza

- La nobleza sin duda interesa particularmente el lector

Lemma	Frec. abs. (fragm)	Frec. rel (fragm)	Frec. abs. (incunable)	Razón
caballería	13	0,00134	0,00043	3,13962
caballero	21	0,00217	0,00117	1,85388
noble	18	0,00186	0,00071	2,61203
nobleza	6	0,00062	0,00017	3,62263
jurar	6	0,00062	0,00008	8,1147
espada	6	0,00062	0,00014	4,31633
caballo	1	0,0001	0,00015	0,69003

- La caballería (¿como institución política tan solo ?) también

Número de marcas por parte y capítulo

Parte	Capítulos	Marcas	Ratio	Tema
I, 1	13	0	0	La bienaventuranza del príncipe
I, 2	34	27	0,79412	Las virtudes
I, 3	11	4	0,36364	Las pasiones
I, 4	7	23	3,28571	Las costumbres de los reyes
II, 1	24	9	0,375	Las mugeres
II, 2	22	9	0,40909	Los fijos
II, 3	20	18	0,9	Los siervos
III, 1	20	8	0,4	La opinión de los filósofos sobre la Ciudad
III, 2	36	32	0,88889	El gobierno en tiempos de paz
III, 3	23	21	0,91304	El gobierno en tiempos de guerra
Total	210	151	0,80749	

Parte I, 4 : « De las costumbres de los reyes »

- 24 marcas en total
- Un apego a la nobleza ante todo y a la defensa de la idea de superioridad de la nobleza (16 marcas)

buenaſ ſpanas do qen bque calos q bque
apartados ſon fiudos i mottenes ſes i por el a
tempo los nobles por la ce pama qan colo 3
guenos ſalle muñ enſenados i muy arce ſes
i fiefas costumbres ſon muchos de alabas elos
en ellis ſe muchos mas elos tanta q deneſi

on loſeruos i apreſos i
de ſus adelos fa q ſuſa les actas por elas.
nolleſa qndo mas alogada es al ſu començo
tudo es mas antiqua e mayor q ellos per
papez mas nobles q ſus padres en ſobrepueſa
i despoſtez a los deſcendo auer mayor onſraſ
i ſu enemis ſu enemis no don

ro q los nobles ſon loſy ro i ſuſa la pueras
lo prouar por eſtas doſ ſafonos la pueras
es qſia altega i la onya leuantan el entedimy
endo de aq. qſia hā apeñar i aſtudiar en como
la guadara i en como la aqꝫetara i como por
el etudio ſe gane la ſtileſa i la ſabiduria ſegun

to aqyl i ſin nobleſa mas facē al ome meſmo
qſia aventureado ca el filoſofo diſe q el tal pia
es bié aventureado fin ſefo oes ſin ſefo aue aqyo
do i qere deſir q ſe maſ al ogo dia aca q no
ſabe bié uſar de las maſias i aſies maſ ſin ſefo
onde la nobleſa i el poderio aqyl muy bié ſe
acompanan oſias maſias ca los q ſon maſ i no

Parte II, 3 : « El gobierno de los siervos »

- Por estudiar !

1 Introducción

2 Métodos de extracción de los fragmentos anotados

3 Estudio lexicométrico y primeros resultados

4 Conclusiones

Un programa de lectura coherente

- Una lectura política clara e intereses de lectura que se pueden identificar con relativa seguridad
- Estos intereses se inscriben bastante bien en el marco de los debates políticos del siglo 15, sobre la nobleza en particular.

Por hacer

- Completar los análisis de frecuencia con el contexto (co-ocurrencias)
- Ampliar el estudio a los otros tipos de marcas y a otros manuscritos
- Añadir una fase de corrección del texto

Preguntas para el futuro

- Confirman estas observaciones las marcas verticales, probablemente producidas por la misma mano ?
- Podemos precisar la localización de las marcas, en la glosa o en la traducción ?
- Podemos ampliar el estudio a otros manuscritos para comparar los intereses de lectura ?

Gracias

Gracias por la atención!

Références I

- [1] Isabel ACERO-DURÁNTEZ, María Jesús DÍEZ GARRETAS et José Manuel FRADEJAS RUEDA. « Las Versiones A y B de La Traducción Castellana Del De Regimine Principum de Gil de Roma ». *Actas Del IX Congreso Internacional de La Asociación Hispánica de Literatura Medieval : A Coruña, 18-22 de Septiembre de 2001*. Sous la dir. de Mercedes PAMPÍN BARRAL et Carmen PARRILLA GARCÍA. T. I. 2005, p. 227-233. ISBN : 84-96259-72-2. URL : <http://www.ahlm.es/IndicesActas/ActasPdf/Actas9.1/09.pdf>.
- [2] Thibault CLÉRICE. *You Actually Look Twice At It (YALTAi) : Using an Object Detection Approach Instead of Region Segmentation within the Kraken Engine*. 2022. preprint.
- [3] Thibault CLÉRICE, Pinche ARIANE, Vlachou-Efstathiou MALAMATENIA, Chagué ALIX, Camps JEAN-BAPTISTE, Gille-Levenson MATTHIAS, Brisville-Fertin OLIVIER, Fischer FRANZ, Gervers MICHAELS et Boutreux AGNÈS. « CATMuS Medieval : A Multilingual Large-Scale Cross-Century Dataset in Latin Script for Handwritten Text Recognition and Beyond ». (2024). URL : <https://inria.hal.science/hal-04453952/> (visité le 14/03/2024).
- [4] Matthias GILLE LEVENSON. « Le Regimiento de Los Príncipes et sa glose : étude et édition numérique de la partie sur le Gouvernement de la cité en temps de guerre (III, 3) ». Codir. Carlos HEUSCH et Jesús R. VELASCO. École Normale Supérieure de Lyon, 2023. URL : <https://theses.hal.science/tel-04337406>.
- [5] Benjamin KISSLING. « Kraken - an Universal Text Recognizer for the Humanities ». DH2019 : Complexity. Utrecht, 2019. URL : <https://dev.clariah.nl/files/dh2019/boa/0673.html>.
- [6] Franco MORETTI. *Distant Reading*. Verso, 2013.
- [7] Noëlle-Laetitia PERRET. *Les traductions françaises du ‘De regimine principum’ de Gilles de Rome : Parcours matériel, culturel et intellectuel d’un discours sur l’éducation*. Education and Society in the Middle Ages and Renaissance 39. Leiden ; Boston : Brill, 2011. ISBN : 90-04-20619-1.
- [8] Joseph REDMON, Santosh DIVVALA, Ross GIRSHICK et Ali FARHADI. « You Only Look Once : Unified, Real-Time Object Detection ». *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016, p. 779-788. URL : https://www.cv-foundation.org/openaccess/content_cvpr_2016/html/Redmon_You_Only_Look_CVPR_2016_paper.html (visité le 14/03/2024).